DOCUMENT RESUME

ED 140 665 FL 008 667

AUTHOR Teitelbaum, Herta
TITLE The Validity of Various Techniques Measuring Children's Bilingualism. Working Papers on Bilingualism, No. 13.
INSTITUTION Ontario Inst. for Studies in Education, Toronto. Bilingual Education Project.
PUB DATE May 77
NOTE 34p.; For related documents, see FL 008 663-669
AVAILABLE FROM Bilingual Education Project, The Ontario Institute for Studies in Education, 252 Bloor St. West, Toronto, Ontario, Canada M5S 1V6 (as long as supply lasts)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Bilingual Education; *Bilingualism; *Bilingual Students; Elementary Education; English (Second Language); *Language Proficiency; Language Skills; Language Tests; *Language Usage; *Measurement Techniques; *Predictive Validity; Predictor Variables; Rating Scales; Self Evaluation; Spanish; Spanish Speaking; Speech Communication; Student Evaluation; Teacher Attitudes; Test Validity

ABSTRACT

Data on language competence and language use were obtained for 99 Spanish/English bilingual elementary school children (kindergarten through fourth grade) from Albuquerque. The children's language skills were assessed by teachers, interviewers, and the children themselves. The teachers and the children also estimated the extent to which the children used Spanish and English in various settings of the home, the neighborhood and the school. As more objective measures of the children's language skills, they performed three tasks--word naming, sentence repetition, and free speech. The relationships among the various measurement techniques were analyzed. The three tasks were well interrelated and were thought to measure the same skills. The relationship between language use and language skill ratings was generally not very strong, both in the case of the children and the teachers. Only a few of the language use variables were found useful in the prediction of the children's performance on the tasks. Of the language skill ratings, the interviewer judgments were the best predictors of the tasks. The teacher ratings were moderately well correlated with the performance measures, and the children's self-ratings least so. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED140665

The Validity of Various Techniques Measuring Children's Bilingualism

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Herta Teitelbaum[1]
University of New Mexico

Abstract

Data on language competence and language use were obtained for 99 Spanish/English bilingual elementary school children (kindergarten through fourth grade) from Albuquerque. The children's language skills were assessed by teachers, interviewers, and the children themselves. The teachers and the children also estimated the extent to which the children used Spanish and English in various settings of the home, the neighborhood and the school. As more objective measures of the children's language skills, they performed three tasks--word naming, sentence repetition, and free speech. The relationships among the various measurement techniques were analyzed. The three tasks were well interrelated and were thought to measure the same skills. The relationship between language use and language skill ratings was generally not very strong, both in the case of the children and the teachers. Only a few of the language use variables were found useful in the prediction of the children's performance on the tasks. Of the language skill ratings, the interviewer judgments were the best predictors of the tasks. The teacher ratings were moderately well correlated with the performance measures and the children's self-ratings least so.

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Working Papers on Bilingualism
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

FL008667

The Validity of Various Techniques Measuring Children's Bilingualism

Herta Teitelbaum
University of New Mexico

Introduction

This paper is concerned with the analysis of the adequacy of instruments used to measure the sociolinguistic patterns of bilingual (Spanish/English) elementary school children in Albuquerque.

The most widely employed methods of collecting sociolinguistic data from bilinguals have been questionnaires and interviews. They typically include items asking bilinguals to rate themselves on a given scale in terms of language skills (reading, writing, understanding, and speaking) and language use in various settings. It is not surprising that questions of reliability and validity have, in the past, principally focussed on these measures. There are obvious advantages to collecting data by questionnaire or interview technique. Direct observation by field workers is often too time-consuming and too expensive. On the other hand, serious problems may be associated with self-reported data, and it can be said that the validity of language background ratings, i.e. of self-reports of language proficiency and language use, has by no means been established.

Studies relating self-reports of language background by children to ratings by others, such as interviewers or teachers, seem to have obtained reasonably high correlation coefficients. Hoffman (1934) reports high correlations between language background data collected from bilingual children and (a) interviewer ratings based upon visits with the children's parents (r=.82), and (b) ratings of a school principal acquainted with the language background of the children (r=.77). Arsenian (1937) claimed additional evidence for the validity of Hoffman's bilingual background questionnaire on the basis of comparison between bilingual children born in the United States and children born abroad. The foreign-born children rated themselves significantly higher on use of the ethnic language than did the native-born children (p. 59). However, attempts to correlate self-reported language background data with more objective measures such as tasks designed to test the bilinguals' language

skills have been less successful (Macnamara, 1969; Fishman & Terry, 1971). For example, in Fishman's study of Puerto Rican bilinguals of Jersey City, correlations between reported language background data and the performance criteria (word naming, word association, listening comprehension) were always below .50 (p. 188).

Language use self-ratings and language competence self-ratings have typically been combined into a single score representing a subject's degree of bilingualism (Hoffman, 1934; Arsenian, 1937). This technique is useful and appropriate if the two sets of variables are in fact closely associated; that is, if proficiency closely indicates level of use and level of use closely indicates proficiency. However, in studies in which language use and language skill ratings have been considered separately, they were found to be differentially related to other measures of bilingualism. Macnamara (1969), in a study of French/English bilingual children in Montreal, found language skill ratings to be better predictors of the children's performance on various language tests than self-ratings on language use (pp. 80-97). In Fishman's study of Puerto Rican bilinguals, self-ratings of language use and language skills were observed to be factorially separate dimensions (Fishman & Terry, 1971, p. 183). These results suggest that the presumed relationship between language use and language competence of bilinguals may not be always as close as is supposed. Accordingly, combining both kinds of ratings does not appear to be warranted without initial investigation into the extent of their interdependence.

A further problem in the validation of measurement techniques concerns the seemingly differential efficacy of various language tasks for ascertaining the extent of bilingual skills. Fishman and Terry (1971) found word naming to be more highly predictive of self-ratings than word association and listening comprehension tasks. Macnamara (1969), however, found word naming to be less predictive of tests in the four language skills (reading, writing, listening comprehension, and speaking) than other measures such as speed of reading or word completion. These findings indicate that individual tasks may be measuring different aspects of a subject's language skills. It is, therefore, necessary to investigate the consistency with which performance criteria actually measure language skills before conclusions can be drawn concerning the utility of using any one or any combination of these tasks for evaluative purposes.

The majority of studies validating language background self-reports have been carried out with data from

bilinguals nine years or older. The age of children in Hoffman's study (1934) ranged from nine to fifteen, those in Arsenian's investigation (1937) from nine to fourteen. The Puerto Rican bilinguals in Fishman's study (1971) were thirteen years and older. Macnamara (1969) used bilingual sixth graders as subjects in his investigation. Although self-reports on language skills and language use of children under the age of nine have frequently been used in research as an index of these children's bilingualism (Timmins, 1971; Mallory, 1971; Edelman, Cooper & Fishman, 1971; Ortiz, 1975), scarcely any information is available on the validity of self-ratings of very young children.

The data available from our study will, it is hoped, shed further light on some of the yet unresolved issues which were discussed above concerning the validation of methods used in the collection of language background data.

Background of the Study

Setting and Subjects. The investigation was carried out with children of Spanish background in a public elementary school. The school is located in an Albuquerque barrio which claims a 96% Spanish surnamed population and offers a bilingual education program for children from kindergarten through fourth grade. In this program, Spanish is taught as a second language in lessons ranging in length from twenty minutes to one hour daily. During the remainder of the day, Spanish is used informally. The children can be considered to be generally fluent English speakers who vary extensively on their ability to understand and speak Spanish.

The subjects of the study were children who participated in the bilingual education program during the school year 1974-75. This group included students from kindergarten through grade four. Initial interviews were attempted with 145 children. Eight of these 145 children were recent immigrants from Mexico and ultimately were excluded from the population used for analysis. It was felt that these children comprised a special class in the community, and that their lower English and higher Spanish ability and use would bias the data. Thirty-eight more children were excluded because of their lack of ability to follow instructions in Spanish. These exclusions left a total of 99 children for whom adequate data were available for analysis.

Instruments. Data concerning the children's language proficiency and language use were collected from the

teachers, the interviewers, and the children themselves. These data are subjective in origin. For the children, an interview schedule was used to obtain data on (a) their receptive and productive skills in Spanish as well as English, (b) their estimated Spanish and English use in the home, the neighborhood, the school, and in church, and (c) their attitude toward Spanish and English. The questionnaire was adapted from the Spanish/English language dominance assessment schedule developed by Spolsky and Murphy (Spolsky, Murphy, Holm & Ferrel, 1972).

Upon completion of the interview with each child, the interviewer rated the child's comprehension and production skills in Spanish and English. This provided a second estimate of the child's language proficiency.

Teachers from five of the six classes which participated in this study filled out an evaluation form in which they rated the ability of each child in the class to understand and speak both Spanish and English. In addition, the teachers assessed the extent to which Spanish and English were used by each of the children in various school settings and in the home. They were also asked to indicate the first language of the children if they knew it.

Language proficiency was typically rated on a scale ranging from one through four or five. A value of one indicated a judgment that the child did not understand or speak the language in question; three indicated a belief that the skills were adequate; and four or five that they were excellent, in the sense that a native of the same age would have the same level of skills. Language use was rated on a scale from one through five: one meant that only English was used; three that English and Spanish were spoken equally as much; and five that only Spanish was spoken.

In order to assess the children's language skills more objectively, they were asked to carry out three tasks in both Spanish and English--word naming, sentence repetition, and free speech. These tasks will be referred to as performance measures in the text.

In the word naming test which was also adapted from Spolsky and Murphy's language dominance assessment schedule (Spolsky et al., 1972), the children named as many things as they could see or do in the kitchen, the yard, the neighborhood block, and in school. One minute was allowed for each domain. The score for each child was based on the number of words enumerated for each domain.

In the sentence repetition task, the children repeated a series of fifteen stimulus sentences which were given to them on tape. This measure was adapted from a study by Natalicio and Williams (1971), who used sentence repetition for oral language assessment of Black and Mexican-American children. However, the evaluation of the performance of the children on this task differed from that of Natalicio and Williams. First, the number of correctly repeated words in each sentence was coded in this study. Secondly, the fluency with which the repetition was made was judged on a scale from one to five. A value of five indicated that the sentence was repeated immediately and fluently; a value of three was given when the child repeated the sentence haltingly; a one was marked when the child made an attempt but the repeated sentence was incoherent. If the child did not say anything at all, a zero was given.

Finally, in the free speech task, the children were asked to describe a set of three pictures and then were encouraged to talk about them. The pictures were illustrations of an airplane or a firetruck, of animals, and of children playing. Three scoring methods were employed for this task. First, the extent to which English was used in the English task and Spanish in the Spanish task was scored on a scale from one to five. A value of five conveyed that Spanish only was used, a three that Spanish and English were used equally, and a one that only English was spoken. The child was also judged on effort expended during this task. The effort scale had also five values. A value of one meant that a great deal of effort was made, a three was given for some effort, and a one for hardly any effort at all. Since the picture task gave the children an opportunity to speak freely, they were furthermore judged on their communicative ability. Again, a scale with five values was used for this purpose. A five was given when the judges thought the child possessed very good command of the language, a three for fair command, and a one for poor command.

Both the sentence repetition and the free speech tasks were judged by two Spanish/English bilinguals. The intercoder reliability for all scales was above .90 as measured by correlational analysis.

The children were interviewed twice--once in Spanish and once in English. Both the language background data and the three language proficiency tasks from the two interviews were recorded on tape. Half of the interviews were first given in Spanish and half first in English. A copy of both the teacher evaluation form and the children's interview schedule as well as a list of the Spanish and

English sentences of the sentence repetition task can be found in the Appendix.

Interviewers. The interviewers were university students who were proficient in both Spanish and English and had, in varying degrees, exposure to and familiarity with sociolinguistics. To establish a positive relationship with the children, the interviewers visited the classroom at least once a week for a period of time to become acquainted with the children. The children were most enthusiastic about the weekly visits of the interviewers and, for the most part, were eager to participate when the interviews began.

Analysis of Relationships among Measurement Techniques

In the investigation of the relationships among the measures of bilingualism used in this study, three aspects were specifically analyzed: the relationship among the performance criteria, the relationship between language ability and language use ratings, and the relationship between subjective ratings (of language skills and language use) and the performance measures.

All children were found to be fluent English speakers, and, therefore, there was no variation among them on their ability to understand and speak English. For this reason, analysis of the measurement techniques is limited to their value for predicting Spanish ability and use, and only the Spanish performance tasks are included in this investigation. To relate various subsets of the data, Pearson product moment correlations were used throughout.

Relationship among Performance Criteria. The purpose of this analysis was to determine how consistently the three tasks--word naming, sentence repetition, and free speech--measured what they were designed to measure, namely the children's skills in Spanish. The intercorrelations of the performance criteria are given in Table 1. It can readily be seen that the measures were highly inter-

---

Insert Table 1 about here

---

correlated. No pattern emerged suggesting a particularly strong relationship for any one combination of the measures. The correlations ranged from .590 to .755 for the relationship between word naming and free speech, from .585 to .648 for the relationship between word naming and sentence

Table 1

INTERCORRELATIONS OF SPANISH PERFORMANCE TASKS

	Word Naming			Free Speech			Sentence Repetition	
Word naming	Yard	Block	School	Span.Use	Effort	Communic. ability	no. of correct words	fluency
Kitchen	.710*	.771*	.718*	.590*	.679*	.708*	.623*	.623*
Yard		.710*	.688*	.637*	.732*	.722*	.628*	.594*
Block			.780*	.627*	.736*	.755*	.625*	.648*
School				.620*	.698*	.739*	.597*	.585*
Free Speech								
Spanish use					.820*	.862*	.598*	.587*
Effort						.936*	.668*	.645*
Communic.ability							.696*	.678*
Sentence Repetition								
No. of correct words								.953*

* $p < .001$

Table 2

FACTOR DERIVED FROM SPANISH PERFORMANCE MEASURES

	Factor
Word naming	
kitchen	.813
yard	.814
block	.859
school	.821
Free speech	
Spanish use	.809
effort	.909
communicative ability	.940
Sentence repetition	
fluency*	.731

* Since the two scoring methods of the sentence repetition task (number of correct words and fluency) were so highly intercorrelated (above .90), only one score, the fluency rating, was included in the factor analysis.

repetition, and from .587 to .696 between free speech and sentence repetition. Children who score high on one task also tend to score high on the other two tasks. It can be concluded that all three tasks measure equally well the children's Spanish skills. To obtain further proof of the concurrent validity (Anastasi, 1976, pp. 140-141) of the three tasks, they were factor analyzed. The findings of the factor analysis are reported in Table 2. As expected, only one factor emerged, with loadings from the variables

---

Insert Table 2 about here

---

ranging in magnitude between .731 and .940. This finding offered additional support in the claim that all tasks were indeed assessing the same skills.

Although the findings in this study may, on the surface, appear to be at odds with those of Macnamara (1969) or Fishman & Terry (1971) who observed some tasks to be more useful in the prediction of bilingualism than others, this may in fact not be the case since both the focus and the procedures of this study differed from the others. Macnamara and Fishman used difference scores in their analyses: that is, for each task, the scores obtained when carrying out the task in one language were subtracted from the scores obtained when the task was performed in the second language. This technique produced a score which represented a subject's language dominance. In our study, however, it was immediately determined that the children were dominant in English and did not vary on this dimension; we thereafter focused solely on their Spanish skills and their degree of Spanish use. Therefore, the findings of Fishman and Macnamara are not, strictly speaking, comparable to ours and do not impeach our conclusion that, for the children in our study, all three tasks used are equally appropriate instruments for assessing their Spanish skills.

Relationship between language ability ratings and language use ratings. It was argued that ratings of language skills and language use should not be combined into a single score presumed to represent a subject's degree of bilingualism unless the two kinds of variables are highly predictive of one another. If the relationship is only a weak one, language use and language ability ratings should be regarded as separate aspects of bilingualism.

In this study, estimates of the children's Spanish competence and their use of Spanish and English in various

settings of the school and the home were obtained from the teachers and the children themselves. It should not be unreasonable to assume that children who rate themselves high on Spanish skills also rate themselves high on use of Spanish. Similarly, teachers who give the children high ratings on Spanish competence would be expected also to give them high ratings on Spanish use. To test this hypothesis, it will be investigated, for teachers and children separately, how strong the association is between the Spanish proficiency ratings and the language use ratings.

The teacher ratings of the children's Spanish skills were related to the teacher estimates of the children's language use in several school settings (in class, on the playground, and in interaction with the teacher) and in the home. The correlations are given in Table 3 and suggest that the teacher judgments of Spanish competence

---

Insert Table 3 about here

---

are moderately well related to the ratings of Spanish use, particularly in the school situations. The coefficients for the school settings range between .424 and .652. The association between ratings of Spanish skills and estimated Spanish use in the home is not as strong (.329 and .477). This might suggest that the teachers were less familiar with and more likely to guess on the children's use of Spanish in the home.

The coefficients do not vary greatly for correlations between production ratings and language use items, on the one hand, and between comprehension ratings and language use evaluations, on the other, because the teachers did not distinguish, to any great extent, between receptive and productive Spanish skills in their judgments.

The children's self-ratings of Spanish proficiency were correlated with their estimation of (a) Spanish use overall and with the teacher, (b) language use in the home with parents and siblings, and (c) language use in the community--with neighbors, friends, and in church. The coefficients are presented in Table 4 and reveal that

---

Insert Table 4 about here

Table 3

CORRELATIONS BETWEEN TEACHER RATINGS OF SPANISH PROFICIENCY AND TEACHER RATINGS OF LANGUAGE USE

Language use ratings	Spanish proficiency ratings Production	Comprehension
School domain		
in class	.424*	.459*
on playground	.481*	.505*
response to Spanish question by teacher	.652*	.580*
initiating conversat. in Spanish with teacher	.632*	.632*
Home domain	.477*	.329*

* $p < .01$

Table 4

CORRELATIONS BETWEEN CHILDREN'S SELF-RATINGS OF SPANISH PROFICIENCY AND CHILDREN'S SELF-RATINGS OF LANGUAGE USE

Language use self-ratings	Spanish proficiency self-ratings Production	Comprehension
Spanish use overall	.457*	.344*
Spanish use w. teacher	.414*	.290**
Home domain		
child to parents	.335*	-
parents to child	-	.364*
child to siblings		.213**
siblings to child	-	-
Neighborhood domain		
neighbors	-	.250**
friends	-	-
church	-	-

* $p < .01$
** $p < .05$

the self-ratings of Spanish skills are, in general, not very strongly related to the self-ratings on Spanish use. The highest coefficients were obtained for Spanish use overall and Spanish use with the teacher. Even there the correlations were always below .50. Both language use items were more strongly related to self-ratings of Spanish production than to Spanish comprehension ratings. Reports of language use at home were only poorly associated with Spanish proficiency ratings. The highest correlations were obtained for Spanish production ratings and Spanish use by the child when speaking to parents (.335) and for ratings of Spanish use by parents when speaking to the child and self-ratings of Spanish comprehension (.364). Language use ratings for interactions with siblings, friends, neighbors, and in church are almost completely unrelated to professed Spanish skills.

The teacher ratings were, it is true, found to be more highly interrelated than were those of the children; nevertheless, the coefficients in neither case were sufficiently high to warrant the combination of both types of ratings into a single score. Moreover, combination could have masked relationships that could be important for evaulative purposes. For teachers, language use ratings in school settings, particularly those addressed to interaction between teacher and child, were more closely related to Spanish proficiency ratings than ratings of language use in the children's home. For the children, it is particularly interesting to note that of all self-reported language use items, ratings of overall Spanish use best predicts Spanish proficiency self-ratings. Questions which require the children to estimate the extent to which Spanish is used in specific situations were less well, and sometimes not at all, related to self-ratings of Spanish skills than the question which asked the children to indicate more globally the extent to which they spoke Spanish. Had we combined all language use ratings, these distinctions would surely have been obscured.

Relationship between subjective ratings (of language skills and language use) and the performance criteria. It was determined above that no strong association exists between language use and language ability ratings obtained from both the children and the teachers. Therefore, we also expect these two kinds of variables to be differentially related to the tasks. Our expectations were borne out. The coefficients were generally stronger and more consistent between performance criteria and proficiency ratings than between performance measures and language use ratings. These findings confirm those of Macnamara (1969) who also observed that for the children in his study

self-ratings of language skills better predicted performance on language tasks than did self-ratings of language use.

In the analysis of the relationship between ratings of Spanish skills and the performance criteria, the ratings of the children's Spanish proficiency by the interviewers, by the teachers, and by the children themselves were related to the tasks. From Table 5 it is apparent that the

---

Insert Table 5 about here

---

interviewer ratings were the best predictors of the children's performance on the three tasks. The correlations for the interviewer assessments of Spanish production skills and the tasks ranged between .660 and .788. The teacher ratings of both comprehension and production were moderately useful in the prediction of the criteria with correlations ranging between a low of .269 and a high of .445. Since the teachers did not differentiate greatly between comprehension and production ability of the children, no sizeable differences in the coefficients for performance tasks and production skills, on the one hand, and the performance measures and the comprehension skills, on the other, were observed. The children's self-ratings of Spanish skills were least predictive of performance of the children on the three tasks. Coefficients were always below .40. The correlations were somewhat higher between Spanish comprehension self-ratings and the performance measures than between the production ratings and the performance criteria.

No evidence was found to suggest that the interviewer ratings were particularly strongly related to any one task. The teacher ratings were somewhat more closely associated with the free speech task than with the other two measures. The children's self-ratings were, to some extent, more predictable from word naming than from sentence repetition and free speech.

Besides investigating the overall relationship between language use ratings and the performance criteria, it will be analyzed specifically how reported Spanish use in particular settings (such as the school or the home) is related to the corresponding domain of the word naming task. One might, for example, suppose that if a teacher reports that a child speaks more Spanish than English in the home, the child will obtain a relatively high score on

Table 5

CORRELATIONS BETWEEN SPANISH PROFICIENCY RATINGS (BY INTERVIEWERS TEACHERS, AND CHILRDEN) AND SPANISH PERFORMANCE MEASURES (TOTAL)

Spanish performance tasks	Spanish Proficiency Ratings					
	Interviewers		Teachers		Children	
	Production	Comprehension	Production	Comprehension	Production	Comprehension
Word naming						
kitchen	.719*	.420*	.269*	.323*	.173	.345*
yard	.737*	.424*	.303*	.311*	.237**	.339*
block	.691*	.389*	.317*	.341*	.145	.244*
school	.788*	.427*	.349*	.368*	.222**	.360*
Free speech						
Spanish use	.701*	.277*	.427*	.376*	.194**	.218**
effort	.760*	.359*	.435*	.445*	.242**	.249*
communicat.ability	.779*	.381*	.428*	.430*	.247**	.267*
Sentence repetition						
no of correct words	.677*	.471*	.404*	.358*	.186	.264*
fluency	.660*	.425*	.379*	.306*	.193**	.214**

* $p \leq .01$
** $p \leq .05$

Table 6

CORRELATIONS BETWEEN LANGUAGE USE RATINGS BY TEACHERS AND SPANISH PERFORMANCE TASKS

LANGUAGE USE RATINGS BY TEACHERS	Word Naming				Free Speech			Sentence Repetition	
	Kit-chen	Yard	Block	School	Span. use	Effort	Communic-ability	No. of correct words	fluency
SCHOOL DOMAIN									
in class					.183*				
on playground					.207*				
response to Span. quest. by teacher	.304**	.273**	.196*	.292**	.278**	.271**	.278**	.210*	
initiating convers. in Spanish with teacher	.328**	.288**	.295**	.365**	.324**	.381**	.417**	.253**	.189*
HOME DOMAIN									

* $p < .05$
** $p < .01$

the kitchen portion of the word naming task. Similarly, if a child reports speaking a lot of Spanish in school, we would expect a relatively higher score in the school domain of word naming than from a child who reports only speaking English in school.

The correlations reflecting the relationship between the teacher ratings of the children's language use and the children's performance on the three tasks which are reported in Table 6 are disappointingly low. Only the ratings

---

Insert Table 6 about here

---

of Spanish use by children in conversation with the teacher show some positive relationship to the tasks. The coefficients range between .189 and .381. Almost all other language use items appear unrelated to task performance. There was no indication that language use ratings by teachers for the school settings were particularly strongly associated with the school portion of word naming.

In Table 7 the correlations for the children's self-ratings on language use and the performance criteria are

---

Insert Table 7 about here

---

presented. Of all language use variables, ratings of overall Spanish use are moderately well related to the performance criteria. The coefficients range from a low of .401 to a high of .577.

The coefficients for reported use of Spanish with the teacher and the tasks are generally lower (between .257 and .386). It was also observed that the ratings of Spanish use with the teacher were more strongly associated with the school portion of word naming than with the other domains. Ratings of Spanish use with parents and siblings were only poorly correlated with the performance criteria. Estimated Spanish use by children with their parents is the only variable which was found to be consistently related to the tasks; however, the correlations were generally low, ranging from .203 to .367. It was not noted that language use ratings at home were particularly strongly associated with the kitchen or the yard portions of word

Table 7

CORRELATIONS BETWEEN LANGUAGE USE RATINGS BY CHILDREN AND SPANISH PERFORMANCE TASKS

LANGUAGE RATINGS BY CHILDREN	Word Naming				Free Speech			Sentence Repetition	
	kit-chen	yard	block	school	Span. use	effort	communic ability	no. of correct words	fluency
Span.use overall	.486**	.448**	.475**	.577**	.401**	.557**	.532**	.444**	.403**
Span.use with teacher	.297**	.257**	.324**	.386**	.383**	.344**	.381**	.287**	.301**
Home Domain									
child w/parents	.276*	.203	.295*	.354**	.352**	.298*	.367**	.257*	.253*
parents w/child							.231*	.231*	
child w/siblings	.264*			.228*				.210	
siblings w/child	.254*		.233*	.240*					
Neighborhood Dom.									
neighbors									
friends		.211							
church									

* $p < .05$
** $p < .01$

naming. Self-reported Spanish use in the community was least useful in the prediction of performance on the three tasks. These correlations were generally close to zero.

Conclusion

It is increasingly important, for purposes of establishing and funding bilingual programs, to determine with greater accuracy the language skills and language dominance of bilingual children. It is not surprising, therefore, that personnel involved in bilingual education both at the elementary school level as well as at universities are deeply concerned with the development and research of instruments which measure more adequately and more efficiently the degree of children's bilingualism. The traditional techniques of assessing bilinguality are being questioned from a number of directions. Zirkel (1976) squarely presents the difficulties with such indices of bilingualism as Spanish surnames, self-ratings or ratings by teachers. The findings of this study confirm the imprecisions incurred in relying on these indices. Although the majority of the children here were of Spanish background and had Spanish surnames, they were all fluent English speakers and used English as the language of communication in most settings. To use the Spanish surname as evidence of these children's Spanish skills would have been a serious misrepresentation of reality. Zirkel (1976) notes that teacher ratings can be "affected by the teacher-rater's competence in each language, by attitudes toward the pupil's native language and culture, and by feelings of job security" (p. 325). He also believes that self-ratings of language dominance by elementary school children is "a questionable procedure" (p. 325). This investigation of the usefulness of reported data on the children's language background by the teachers and by the children themselves had also demonstrated the difficulty of obtaining reliable and valid data. Only a few of the language use ratings by the teachers as well as by the children were found to be related to the three Spanish tasks. Although assessments by the five teachers of the children's Spanish skills were more highly predictive of the children's performance on the tasks than the children's self-ratings of Spanish competence, the correlations were never very high. In a separate analysis, the relationship between teacher ratings and task performance for the individual teachers were analyzed. Substantial variation was observed among the teachers in the prediction of the children's performance from the ratings. It may be that the teachers are generally unfamiliar with the language background of the children and may be likely to judge the children on the

basis of impressionistic rather than realistic assessment. An alternative or complementary explanation for the differences may lie in differing attitudes toward bilingual education itself. This would not be at all surprising in view of the highly politicized situation in the Southwest. Bilingual education is still a hotly debated issue with respect to its goals and consequences. Teachers who strongly support the notion of maintaining the language and cultural heritage of the people of Spanish background may be inclined to overestimate the children's Spanish skills and use. Other teachers may, for a variety of reasons, have less positive feelings toward bilingual education. They may, consequently, underestimate the children's Spanish ability and use.

The children's self-ratings of Spanish skills were the weakest predictors of their performance on the tasks. Again, we can only speculate about the reasons for this finding. Children at this age may not have the sophistication to monitor their language ability, a difficulty which became apparent in interviews with the youngest children who were sometimes unable to indicate the language in which they were speaking at that particular moment. The children too may be biased in estimating their Spanish skills because of attitudes held toward their background. One of the questions the children were asked was whether they liked to speak Spanish and English. Some of them clearly stated their preference for English. Those children might underestimate their actual Spanish skills while others who indicated a strong liking for Spanish may have overestimated their actual competence.

Clearly, the best predictors of the children's performance on the Spanish tasks were the ratings by the trained interviewers. While outside ratings may yield less biased estimates of children's language skills, this procedure may be undesirable from a practical viewpoint because of the additional time it takes for the interviewers to establish rapport with the children in order to secure realistic language data.

Because of the multitude of problems with subjective ratings of children's bilingualism, efforts are increasingly concentrating on developing more objective measures, that is on tests which are designed to measure functionally comprehension and production skills of bilinguals (Zirkel, 1976). However, it cannot be assumed that more objective instruments will solve all problems and provide reliable and valid language data. It is generally recognized that a variety of factors may influence the performance of children on language tests. Children may be affected by the formality of the setting and of the language used in

the testing situation (Hymes, 1971; Labov, 1972). The test content itself may be linguistically or culturally inappropriate for the children taking the test as Brière (1973) has demonstrated. Since the results of language dominance tests will be used in the planning of bilingual education programs in general and in the planning of children's educational curricula in particular, it seems that serious efforts are necessary to ensure that the language data obtained from the children reflect as true and as reliable a picture as possible of their actual language skills. It follows that we should not rely on a single index of degree of bilingualism (e.g. a single language test or a single subjective judgment) but rather that we should develop a comprehensive plan to collect language use and language competency data from several sources. This was done in the study reported here where the subjective ratings and the performance tasks could be cross-validated.

Note

[1]This article is part of a dissertation which was completed in fall 1976. The author is indebted to Dr. Vera John-Steiner, Dr. Bernard Spolsky, and Dr. John W. Oller for their many helpful comments during the preparation of this report. A subvention from the Graduate Student Association of the University of New Mexico in support of this paper is also gratefully acknowledged.

References

Abrahams, R.D., and Troike, R.C. (Eds.). 1972. Language and cultural diversity in American education. Englewood Cliffs, N.J.: Prentice-Hall.

Anastasi, A. 1976. Psychological Testing. New York: MacMillan.

Arsenian, S. 1937. Bilingualism and mental development; a study of the intelligence and the social background of bilingual children in New York City. New York: Teachers College, Columbia University.

Brière, E.J. 1973. "Cross-cultural biases in language testing." In J.W. Oller and Richards, J.C. (Eds.). Focus on the learner, Rowley, Mass.: Newbury House.

Edelman, M., Cooper, R.L., and Fishman, J.A. 1971. "Young Puerto Rican schoolchildren." In J.A. Fishman, R.L. Cooper, R. Ma, et al. Bilingualism in the barrio. Bloomington, Ind.: Indiana University.

Fishman, J.A.; Cooper, R.L.; Ma, R.; et al. 1971. Bilingualism in the barrio. Bloomington, Ind.: Indiana University.

Fishman, J.A., and Terry, C. 1971. "The contrastive validity of census data on bilingualism in a Puerto Rican neighborhood." In J.A. Fishman, R.L. Cooper, R. Ma, et al. Bilingualism in the barrio. Bloomington, Ind.: Indiana University.

Hoffman, M.N.H. 1934. The measurement of bilingual background. New York: Teachers College, Columbia University.

Hymes, D. 1971. "On linguistic theory, communicative competence, and the education of disadvantaged children." In M.L. Wax, S. Diamond, and F.O. Gearing (Eds.). Anthropological perspectives on education. New York: Classic Books.

Kelly, L.G. (Ed.). 1969. Description and measurement of bilingualism. Toronto: University of Toronto.

Labov, W. 1972. "The logic of nonstandard English." In R.D. Abrahams and R.C. Troike (Eds.). Language and cultural diversity in American education. Englewood Cliffs, N.J.: Prentice-Hall.

Macnamara, J. 1969. "How can one measure the extent of a person's bilingual proficiency." In L.G. Kelly (Ed.). Description and measurement of bilingualism. Toronto: University of Toronto.

Mallory, G.G. 1971. "Sociolinguistic considerations for bilingual education in an Albuquerque community undergoing language shift." Unpublished doctoral dissertation. University of New Mexico.

Natalicio, D.C., and Williams, F. 1971. "Repetition as an oral language assessment technique." Austin, Tex.: University of Texas. (ERIC Document Reproduction Service No. ED 051 680).

Oller, J.W., and Richards, J.C. (Eds.). 1973. Focus on the learner. Rowley, Mass.: Newbury House.

Ortiz, L.E. 1975. "A sociolinguistic study of language maintenance in the northern New Mexico community of Arroyo Seco." Unpublished doctoral dissertation. University of New Mexico.

Spolsky, B., Murphy, P., Holm W., and Ferrel A. 1972. "Three functional tests of oral proficiency." TESOL Quarterly, 6, 221-235.

Timmins, K.M. 1971. "An investigation of the relative bilingualism of Spanish surnamed children in an elementary school in Albuquerque." Unpublished doctoral dissertation. University of New Mexico.

Wax, M.L., Diamond, S., and Gearing, F.O. (Eds.). 1971. Anthropological perspectives on education. New York: Basic Books.

Zirkel, P.A. 1976. "The why's and ways of testing bilinguality before teaching bilingually." Elementary School Journal, 6, 323-329.

Appendix

(The English version of the interview schedule and other instruments can be obtained from the author -- Eds.)

TEACHER EVALUATION FORM

CHILD'S NAME _____ DATE _____

GRADE _____ CHILD'S FIRST LANGUAGE

DATE OF ADMISSION TO SCHOOL _____ _____

WHERE WAS THE CHILD RAISED? Martineztown _____ Elsewhere in Albuquerque _____ Other (please specify) _____

LANGUAGE ABILITY PROFILE

Place a check mark in the appropriate column. In your best judgment, how well does the child:

	1. Not at all	2. Poorly	3. Adequately	4. Very well	5. Very fluently
SPEAK ENGLISH					
UNDERSTAND ENGLISH					
WRITE ENGLISH					
READ ENGLISH					
SPEAK SPANISH					
UNDERSTAND SPANISH					
WRITE SPANISH					
READ SPANISH					

LANGUAGE USE PROFILE

Place a check mark in the appropriate column.

In your best judgment, how much Spanish/English does the child use in the following situations:

	1. English only	2. More Eng. than Span.	3. Span. & Engl. equally	4. More Span. than Eng.	5. Spanish only	6. Don't know
(1) When talking with other Span.speak.children in class						
(2) When talking with other Span.speak.children on playground						
(3) When asked a question in Spanish by a teacher						
(4) When initiating a conversation w.a Span.speak.teacher						
(5) At home						

ENTREVISTA GRADE _____ CHILD'S NAME _____

	S	E	Ø
S1. ¿Cómo te llamas?			
S2. ¿Cuántos años tienes?			
S3. ¿Dónde vives?			
S4. ¿Hablas tú en español? Sí___ No___			
¿Cuánto hablas español? Mucho___ Poco___ Poquito___ Nada___			
¿Cómo hablas tú en español? Muy bien___ Bien___ Bastante___ No muy mien___			

(If child is not responding, go to the word naming task. If it still appears that s/he does not understand Spanish, conclude the interview)

S5. ¿Quiénes viven en tu casa contigo? (Mamá, papá, hermanos, abuelos, etc.)

Write down names in rows (a) through (g)	S6."el niño habla" 1 E	2 Es	3 ES	4 Se	5 S	S7. "le habla al niño" 1 E	2 Es	3 ES	4 Se	5 S
(a)										
(b)										
(c)										
(d)										
(e)										
(f)										
(g)										
tus vecinos										
tus amigos en el bloque										
en la iglesia										
otros										

S6. Estamos hablando en español ahora. ¿Hablas español con tu maestra? Sí___ No___
¿Con quién más hablas tú en español? (Mark under "el niño habla")

S7. ¿Quién habla contigo en español? (Mark under "le habla al niño")

S8. ¿A ti te gusta hablar español? Sí___ No___ Ø___
¿Por qué? _____

S9. Cuando la maestra les habla en español, ¿entiendes?
Siempre___ A veces___ Casi nunca___ Ø___

S10. Ademas del español e inglés, ¿hablas otra lengua?
Sí___ No___
¿Cuál lengua? _____ ¿Con quién? _____

English Sentence Repetition

1. The family eats breakfast.
2. Gloria and David drink milk.
3. Daddy goes to work.
4. Mother washes David's neck.
5. She washes his ears.
6. Gloria washes her hair.
7. Gloria cries.
8. Mother helps Gloria.
9. David has a toothbrush.
10. He cleans his teeth with his brush.
11. They are on their knees.
12. The light is not on.
13. Baby has a sock on his leg.
14. David has a brush for his hair.
15. The children wash their hands.
16. They brush their teeth.
17. David gets a little coat.
18. Mother works at home.

Spanish Sentence Repetition

1. Gloria se baña.
2. Gloria tiene sus zapatos.
3. Mamá trabaja en casa.
4. Ella tiene el jabón.
5. Ella tiene jabón en la cabeza.
6. Tiene jabón en la nariz.
7. Gloria llora.
8. Gloria tiene un cepillo de dientes.
9. Se lava los dientes con su cepillo.
10. Están de rodillas.
11. La luz está apagada.
12. Ellos se pueden vestir solos.
13. David puede abotonarse la camisa.
14. El bebé tiene una calcetina en la pierna.
15. Ellos se lavan los dientes.
16. David toma una chaqueta.
17. Ellos van a la escuela hoy.
18. Papá va a trabajar.